



National Aeronautics and Space Administration  
Goddard Earth Science Data Information and  
Services Center (GES DISC)

## README Document for MSAQS02L4

---

A global catalogue of large SO<sub>2</sub> sources and  
emissions derived from the Ozone Monitoring  
Instrument

Goddard Earth Sciences Data and Information Services Center (GES DISC)  
<https://disc.gsfc.nasa.gov>  
NASA Goddard Space Flight Center  
Greenbelt, MD 20771 USA

Last Revised 03/14/2019

**Prepared By:**

Vitali Fioletov, Chris McLinden

Nickolay Krotkov, Can Li, Peter Leonard

---

**Name**

Environment and Climate Change Canada

---

**Name**

NASA/GSFC/614/619

**Reviewed By:**

Nickolay Krotkov

03/14/2019

---

**Reviewer Name**

GES DISC  
GSFC Code 610.2

---

**Date**

**Goddard Space Flight Center  
Greenbelt, Maryland**

## Revision History

---

<i>Revision Date</i>	<i>Changes</i>	<i>Author</i>
5/25/2017	Document Created	Vitali Fioletov
5/30/2017	Document updated	Andrey Savtchenko
6/26/2017	First revision	Nickolay Krotkov
7/10/2017	Second revision	Nickolay Krotkov
3/20/2018	Third revision	Nickolay Krotkov
3/14/2019	Catalog minor version revision 01-04	Andrey Savtchenko

## Table of Contents

<b>1</b>	<b>Introduction .....</b>	<b>5</b>
1.1	Dataset Description.....	5
1.2	Algorithm Background.....	5
1.3	Data Disclaimer.....	6
<b>2</b>	<b>Data Organization.....</b>	<b>7</b>
2.1	HDF5 products .....	7
2.2	File Naming Convention .....	9
2.3	File Format and Structure .....	9
2.4	Key Science Data Fields .....	9
2.5	Science Area .....	10
<b>3</b>	<b>Options for Reading the Data.....</b>	<b>11</b>
3.1	Command Line Utilities .....	11
3.2	Tools for simple visualization and file content view.....	13
3.2.1	HDFView.....	13
3.2.2	Panoply.....	13
3.2.3	Commercial .....	13
<b>4</b>	<b>Data Services .....</b>	<b>13</b>
4.1	GES DISC User Interface.....	14
4.2	Earthdata Search.....	14
4.3	OPeNDAP.....	14
<b>5</b>	<b>More Information.....</b>	<b>15</b>
<b>6</b>	<b>Acknowledgements .....</b>	<b>16</b>
<b>7</b>	<b>References .....</b>	<b>16</b>

# 1 Introduction

---

This document provides basic information about global catalogue of emissions from large SO<sub>2</sub> point sources derived from the Ozone Monitoring Instrument (OMI) on NASA's EOS Aura spacecraft.

Sulfur dioxide (SO<sub>2</sub>) measurements from planetary boundary layer (PBL) SO<sub>2</sub> data produced with the Principal Component Analysis (PCA) algorithm (Li et al., 2013) were used to detect large point emission sources or clusters of sources and estimate annual emissions from them. The sources were identified using different methods, including through OMI measurements utilized in a new emission detection algorithm (Fioletov et al., 2016), and their evolution was traced by estimating annual emissions from each source. The catalogue MSAQSO2L4 file contains the site coordinates, source type, country, source name, annual emissions, annual emission uncertainties, and the number of OMI pixels in the fitting area.

## 1.1 Dataset Description

The catalogue MSAQSO2L4 file contains site coordinates, source type, country, source name, annual emissions, annual emission uncertainties, and the number of pixels in the fitting area.

## 1.2 Algorithm Background

The emission estimates are based on operational OMI planetary boundary layer (PBL) Level 2 SO<sub>2</sub> data produced with the PCA algorithm (Li et al., 2013). For this study, we have scaled constant PCA SO<sub>2</sub> air mass factor (AMF) to a source-specific value based on such parameters as surface reflectivity, solar zenith angle, viewing geometry, surface pressure, cloud fraction, and pressure, and the SO<sub>2</sub> profile shape were also accounted for. As a result, a single site-specific AMF value for each site was calculated (McLinden et al., 2014).

The emission estimate method for linking OMI SO<sub>2</sub> VCDs to SO<sub>2</sub> emissions is based on a fit of OMI VCDs to an empirical plume model developed to describe the SO<sub>2</sub> spatial distribution (as seen by OMI) near emission point sources. The plume model assumes that the SO<sub>2</sub> concentrations emitted from a point source decline exponentially with time and that they are affected by turbulent diffusion that can be described by a two-dimensional (2D) Gaussian function. The total

SO<sub>2</sub> mass is derived from the fit and the emission rate is calculated as the ratio between the total mass and the prescribed lifetime.

The algorithm for emission estimation is described by (Fioletov et al., 2015), the source detection algorithm is available from (McLinden et al., 2016). The original publication of the catalogue is (Fioletov et al., 2016). General information about SO<sub>2</sub> distribution and trends over different regions can be found in (Krotkov et al., 2016) and additional information about volcanic sources is available from (Carn et al., 2017).

Statistical uncertainties (one standard deviation,  $1\sigma$ ) of the annual emission estimates are approximately 10 to 20 kt yr<sup>-1</sup> plus 5 %. The uncertainties caused by the retrieval algorithms including AMF values are estimated at 50–60 %, but comparisons with reliable bottom-up inventories typically indicate agreement to better than 30% (based on the spread of the OMI estimated to reported emissions ratios).

## 1.3 Data Disclaimer

Data should be used with proper citation link:

Vitali Fioletov, Chris McLinden, Nickolay Krotkov, Can Li, Peter Leonard, Joanna Joiner, Simon Carn (2019), Multi-Satellite Air Quality Sulfur Dioxide (SO<sub>2</sub>) Database Long-Term L4 Global V1, Greenbelt, MD, USA, Goddard Earth Science Data and Information Services Center (GES DISC), Accessed [**Data Access Date**] <https://doi.org/10.5067/MEASURES/SO2/DATA403>

The attribution of the sources was done based on our best knowledge and may not always be correct. In some cases, there are several individual sources in close proximity and it is difficult to estimate the contribution of each of them (Fioletov et al., 2017). For others, no definitive information was found on the source origin. Source identification from OMI data is particularly difficult in China, where point sources are numerous and are often located in clusters.

The emission estimation algorithm was developed for point sources. The fitting area size depends on the emission strength. For small sources, is within a 90-km radius, so sources located 180 km away do not affect emission estimates. The algorithm works reasonably well when there are two or even more sources in close proximity (20–30 km) but with no other sources nearby. It is expected that the problem of multiple sources will be resolved in the next versions of the catalogue (Fioletov et al., 2017). Information about such multiple sources can be found in the “Comment” column. There are, however, some regions of China where sources are dense enough that it becomes difficult to apply the algorithm. In these instances, we simply identified hotspots

and included them in the catalogue to have a reasonable representation of the total emissions for such regions. These hotspots are labeled as “area” sources in the catalogue (e.g., Liaoning, Wuan).

## 2 Data Organization

---

### 2.1 HDF5 products

The original catalogue for 2005-2014 in the Microsoft Excel format and its organization is available as the Supplement of the paper: <http://www.atmos-chem-phys.net/16/11497/2016/>

Within the framework of the MEaSURES 2012 SO<sub>2</sub> project, however, these data are converted into HDF5 format. The latter is machine-independent, self-described, contains easy to read attributes, and can be represented as a hierarchy of data objects and folders. In this particular case, the file structure is “flat”, i.e. all data objects are under the “root” level, Figure 1.

Note the File Global Attributes in the “Metadata” tab. They give ample information on the Begin/End data of the data inside the file, the version of the data, the Digital Object Identifier (DOI), etc.

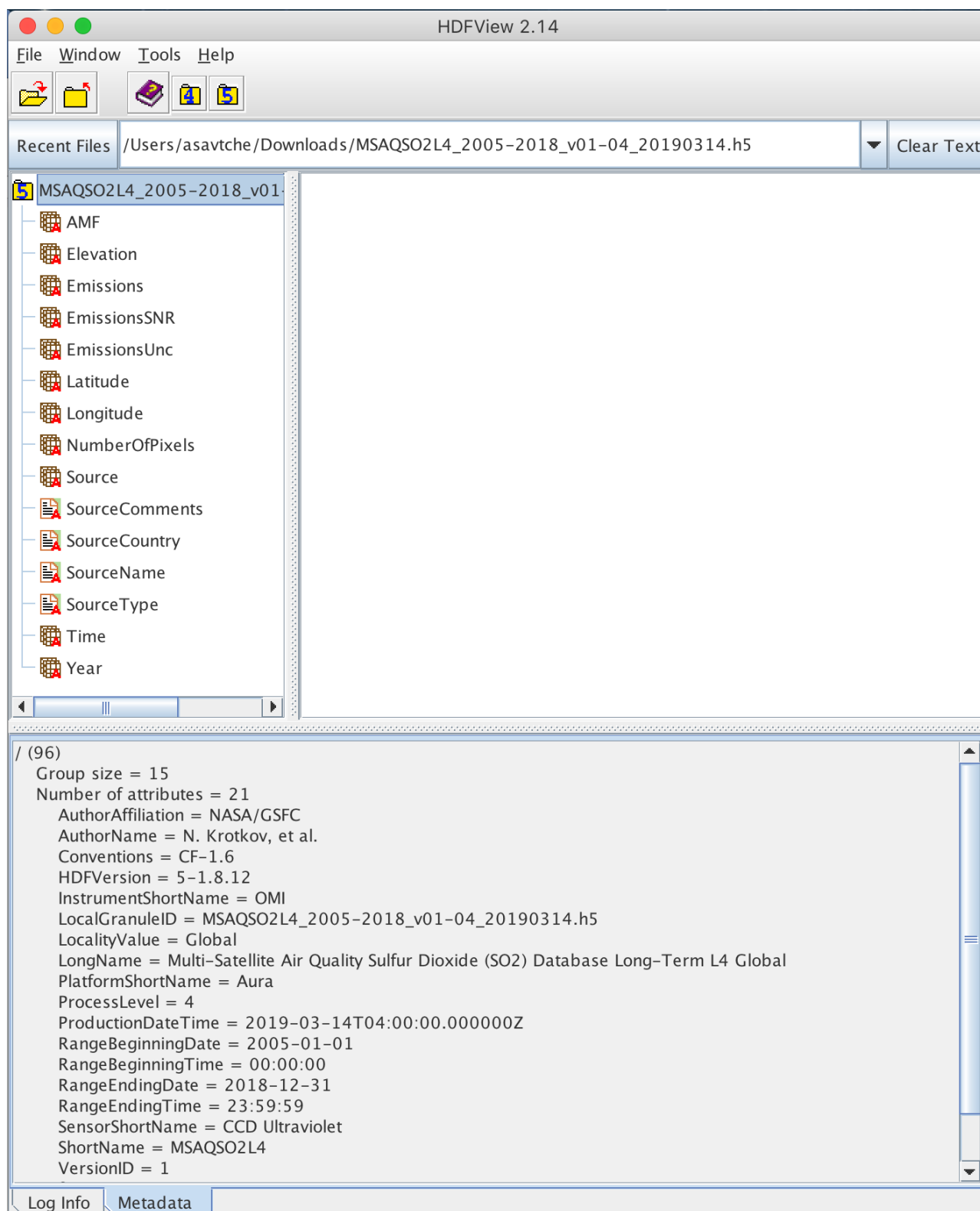


Figure 1. Layout of the MSAQSO2L4 HDF5 file, as represented by the HDFView tool. See ‘Tools for simple visualization, below’. The upper portion shows the data objects in the file, whereas the lower portion are the File Global Attributes.



## 2.2 File Naming Convention

The file naming is fairly transparent:

MSAQSO2L4\_2005-2018\_v01-00.20190314.h5

It says the file is processing Level 4 (L4), the data range is “2005-2018”, data/algorithm version is 1.00, and the file was produced on 2019-03-14.

As new data are added, the file is refreshed, and the file name, as well as the file metadata, the File Global Attributes, are updated to reflect the new content.

## 2.3 File Format and Structure

The MSAQSO2L4 GSSTF3 file is in HDF5 format, which is platform-independent. The structure is very friendly visualized by HDFView as hierarchy of data objects, and file metadata, Figure 1. The key data fields are described below.

The data hierarchy, data dimensions, file and data objects metadata, are much more easily and conveniently described in HDF5 format, and scouted with a viewer like HDFView, rather than if the file was in text format. Conversion to text format is easy using HDFView, or other command-line utilities.

## 2.4 Key Science Data Fields

<i>Column</i>	<i>Content</i>
NUMBER	Site number for internal use (hidden)
LATITUDE	Latitude in deg., negative for the Southern Hemisphere
LONGITUDE	Longitude in deg., negative for the Western Hemisphere
ELEVATION	Site elevation above sea level in m
TYPE	Source type
NAME	Short name for the site identification
COUNTRY	The country of the site location

COMMENT	Additional information about the site. Note that the attribution of the sources was done based on our best knowledge and may not always be correct.
AMF	Air Mass Factor (AMF) values used to adjust emissions based on the site elevation, albedo, etc. These post-processing AMFs can be different from the AMFs assumed in OMI SO <sub>2</sub> retrievals.
EMISSIONS (kt[SO <sub>2</sub> ]/yr) for the year	Estimated emissions for years as indicated in the Metadata
UNCERTAINTY (one standard deviation of the estimated emissions)	One standard deviation of the estimated emissions
RATIOS OF EMISSIONS TO THEIR FITTING UNCERTAINTIES	The ratio of estimated emission to the standard deviation of the estimate
THE NUMBER OF PIXELS	The number of OMI pixels used for the fit for the smallest fitting area (30km by 90 km). Note that larger fitting areas were used for large sources.

## 2.5 Science Area

The catalogue presented herein can be used for verification of traditional (bottom-up) SO<sub>2</sub> emission inventories and identification of missing sources. It can be also used to fill gaps in available inventories, particularly if there are no other sources of information, e.g., for remote volcanoes. Conversely, those sites for which reliable SO<sub>2</sub> emission data are available can be used for OMI SO<sub>2</sub> data product validation. The catalogue could also be used for cross-validation of different satellite data sources, similar to the comparison done for OMI, GOME-2, and SCIAMACHY (Fioletov et al., 2013). This could be particularly useful for cross-validation of new polar orbiting satellite instruments such as the European Commission Copernicus Programme Sentinel 5 precursor (S5p/TROPOMI) and the data from the next- generation geostationary satellites scheduled to be put into orbits over North America (TEMPO), Europe (Sentinel 4), and Asia (Geostationary Environment Monitoring Spectrometer, GEMS) in early 2020s.

## 3 Options for Reading the Data

---

### 3.1 Command Line Utilities

The HDF Group lists a number of HDF5 command line tools on their website:

[http://www.hdfgroup.org/products/hdf5\\_tools/index.html#h5dist](http://www.hdfgroup.org/products/hdf5_tools/index.html#h5dist)

Among them, the one that is the most useful for previewing HDF5 content is **h5dump**. The h5dump is the tool to use to have a very first look at any HDF5-formatted file, even if no documentation describing data is available. The tool by itself does not need extensive documentation, it is very straightforward to use and it's all options can be conveniently listed at the command line using the help option “-h”: h5dump -h

As of the current naming setup of the data fields, a simple list request “-n” will produce the following output, where full paths to groups and individual data fields (datasets) are obvious:

```
h5dump -n MSAQSO2L4_2005-2018_v01-04_20190314.h5
```

```
FILE_CONTENTS {
group    /
dataset  /AMF
dataset  /Elevation
dataset  /Emissions
dataset  /EmissionsSNR
dataset  /EmissionsUnc
dataset  /Latitude
dataset  /Longitude
dataset  /NumberOfPixels
dataset  /Source
dataset  /SourceComments
dataset  /SourceCountry
dataset  /SourceName
dataset  /SourceType
dataset  /Time
dataset  /Year
}
```

We can easily see which years are included, in what order:

```
h5dump -d Year MSAQSO2L4_2005-2018_v01-04_20190314.h5
HDF5 MSAQSO2L4_2005-2018_v01-04_20190314.h5 {
DATASET "Year" {
  DATATYPE H5T_STD_I32LE
  DATASPACE SIMPLE { ( 14 ) / ( 14 ) }
  DATA {
    (0): 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015,
    (11): 2016, 2017, 2018
```

And then dump some data:

```
h5dump -d Emissions MSAQSO2L4_2005-2018_v01-04_20190314.h5
HDF5 MSAQSO2L4_2005-2018_v01-04_20190314.h5 {
DATASET "Emissions" {
  DATATYPE H5T_STD_I32LE
  DATASPACE SIMPLE { ( 518, 14 ) / ( 518, 14 ) }
  DATA {
    (0,0): 103, 101, 91, 138, 131, 79, 112, 109, 132, 142, 227, 29, 134, 72,
    (1,0): 1928, 1764, 2044, 1816, 2302, 2158, 2354, 1806, 2069, 2222, 1771,
    (1,11): 1774, 2102, 1898,
    (2,0): 60, 9, 27, 0, 33, 12, 28, 55, 27, 89, 73, 74, 77, 40,
    (3,0): 58, 66, 28, 8, 68, 0, 34, 37, 67, 23, 153, 129, 92, 44,
    (4,0): 64, 66, 33, 88, 36, 59, 55, 38, 53, 0, 28, 93, 0, 0,
    (5,0): 41, 64, 20, 31, 140, 28, 46, 0, 0, 18, 38, 24, 0, 12,
    (6,0): 0, 22, 27, 6, 1289, 76, 17, 11, 7, 25, 0, 0, 0, 6,
    ,
    . . . .
```

In this example, there are 518 data sources, each source has a maximum of 14 year record.

## 3.2 Tools for simple visualization and file content view

### 3.2.1 HDFView

Among the interactive (with user interface) HDF tools, **HDFView** is one of the simplest to use and install on your local desktop. Strength of HDFView is that it is good for all HDF and netCDF formats. Another strength is that file content is presented graphically in a very friendly fashion – all data groups are presented as folders, where user can easily drill down the file hierarchy, which is especially useful for HDF5. Numerical data sets can be viewed as spreadsheets, and as images (multidimensional datasets). Simple data manipulations are possible. For download and more information, follow this link:

<http://www.hdfgroup.org/hdf-java-html/hdfview/index.html>

### 3.2.2 Panoply

Developed at the Goddard Institute for Space Studies (GISS), the tool is compliant with NetCDF Climate and Forecast (CF) Metadata Convention that is gaining popularity. A strength of the tool is that data can be previewed “remotely” over the network – i.e., user can preview file content of HDF files stored on a remote site, without downloading them. Panoply is available from GISS:<http://www.giss.nasa.gov/tools/panoply/>

With Panoply, data are frequently best viewed through OPeNDAP (more on this service below), by entering the following in the “file”->“open remote catalog” dialog:

<https://measures.gesdisc.eosdis.nasa.gov/opendap/SO2/catalog.xml>

### 3.2.3 Commercial

IDL and MatLAB, are two major commercial data languages, coming with their libraries that fully support all HDF formats. The HDFEOS group has added simple IDL and MatLAB recipes on their web site in particular for GSSTF data in the HDF-EOS5:

[http://hdfEOS.org/zoo/index\\_openGESDISC\\_Examples.php](http://hdfEOS.org/zoo/index_openGESDISC_Examples.php)

## 4 Data Services

---

## 4.1 GES DISC User Interface

GES DISC provides new advanced keyword, spatial, and temporal searches, as well as advanced web services (like OPeNDAP), through its interface. To directly find this particular dataset:

<https://disc.gsfc.nasa.gov/datasets?keywords=MSAQSO2L4>

Users don't have to memorize exact dataset names. It would suffice to enter some relevant keywords, like "SO<sub>2</sub> meas" to find this and other SO<sub>2</sub> datasets, mostly relevant to the MEaSUREs project:

<https://disc.gsfc.nasa.gov/datasets?keywords=so2%20meas>

Also possible:

<https://disc.gsfc.nasa.gov/datasets?keywords=so2%20volcano>

## 4.2 Earthdata Search

Use Earthdata Search Client (EDSC) to find and retrieve data sets across multiple data centers. A shortcut to search for MSAQSO2L4 would be:

<https://search.earthdata.nasa.gov/search?q=MSAQSO2L4>

## 4.3 OPeNDAP

OPeNDAP stands for "Open-source Project for a Network Data Access Protocol". OPeNDAP is a framework that simplifies all aspects of scientific data networking. It provides simple means for parameter and spatial subsetting. The spatial subset is actually a subset by array index - array being the grid data. In the most simplistic case, OPeNDAP can be used to convert data from HDF-EOS5 to ASCII. The data directory hierarchy, of all MEaSUREs-related SO<sub>2</sub> data sets, as served by OPeNDAP, can be viewed in any browser:

<http://measures.gsfc.nasa.gov/opensdap/SO2/contents.html>

## 5 More Information

---

## 6 Acknowledgements

---

We acknowledge the NASA Earth Science Division Atmospheric Composition focus area for funding of OMI SO<sub>2</sub> product development and analysis. The Dutch - Finnish built OMI instrument is part of the NASA's EOS Aura satellite payload. We thank systems engineering, instrument calibration and satellite integration teams for making this mission a success. KNMI and the Netherlands Space Agency (NSO) manage the OMI project. We thank the KNMI OMI team for producing L1B radiance and irradiance data, updating the key calibration data and performing operations together with the U.S. Aura operations team, as well as OMI processing team for continuing support.

## 7 References

- Carn, S. A., Fioletov, V. E., McLinden, C. A., Li, C. and Krotkov, N. A.: A decade of global volcanic SO<sub>2</sub> emissions measured from space, *Sci. Rep.*, 7, 44095, doi:10.1038/srep44095, 2017.
- Fioletov, V. E., McLinden, C. A., Krotkov, N., Yang, K., Loyola, D. G., Valks, P., Theys, N., Van Roozendaal, M., Nowlan, C. R., Chance, K., Liu, X., Lee, C. and Martin, R. V.: Application of OMI, SCIAMACHY, and GOME-2 satellite SO<sub>2</sub> retrievals for detection of large emission sources, *J. Geophys. Res. Atmos.*, 118(19), 11,399-11,418, doi:10.1002/jgrd.50826, 2013.
- Fioletov, V. E., McLinden, C. A., Krotkov, N. A. and Li, C.: Lifetimes and emissions of SO<sub>2</sub> from point sources estimated from OMI, *Geophys. Res. Lett.*, 42, 1–8, doi:10.1002/2015GL063148, 2015.
- Fioletov, V. E., McLinden, C. A., Krotkov, N. A., Li, C., Joiner, J., Theys, N., Carn, S. and Moran, M. D.: A global catalogue of large SO<sub>2</sub> sources and emissions derived from Ozone Monitoring Instrument, *Atmos. Chem. Phys.*, 16, 11497–11519, doi: 10.5194/acp-16-11497-2016, 2016.
- Fioletov, V., McLinden, C. A., Kharol, S. K., Krotkov, N. A., Li, C., Joiner, J., Moran, M. D., Vet, R., Visschedijk, A. J. H., and Denier van der Gon, H. A. C.: Multi-source SO<sub>2</sub> emissions retrievals and consistency of satellite and surface measurements with reported emissions, *Atmos. Chem. Phys. Discuss.*, 17, 12597-12616, 10.5194/acp-2017-485, in review, 2017.
- Krotkov, N. A., McLinden, C. A., Li, C., Lamsal, L. N., Celarier, E. A., Marchenko, S. V., Swartz, W. H., Bucsela, E. J., Joiner, J., Duncan, B. N., Boersma, K. F., Veefkind, J. P., Levelt, P. F., Fioletov, V. E., Dickerson, R. R., He, H., Lu, Z. and Streets, D. G.: Aura OMI observations of regional



- SO<sub>2</sub> and NO<sub>2</sub> pollution changes from 2005 to 2015, *Atmos. Chem. Phys.*, 16(7), 4605–4629, doi:10.5194/acp-16-4605-2016, 2016.
- Li, C., Joiner, J., Krotkov, N. A. and Bhartia, P. K.: A fast and sensitive new satellite SO<sub>2</sub> retrieval algorithm based on principal component analysis: Application to the ozone monitoring instrument, *Geophys. Res. Lett.*, 40(23), 6314–6318, doi:10.1002/2013GL058134, 2013.
- McLinden, C. A., Fioletov, V., Boersma, K. F., Kharol, S. K., Krotkov, N., Lamsal, L., Makar, P. A., Martin, R. V., Veefkind, J. P. and Yang, K.: Improved satellite retrievals of NO<sub>2</sub> and SO<sub>2</sub> over the Canadian oil sands and comparisons with surface measurements, *Atmos. Chem. Phys.*, 14(7), 3637–3656, doi:10.5194/acp-14-3637-2014, 2014.
- McLinden, C. A., Fioletov, V., Shephard, M. W., Krotkov, N., Li, C., Martin, R. V., Moran, M. D. and Joiner, J.: Space-based detection of missing sulfur dioxide sources of global air pollution, *Nat. Geosci.*, 9, 496–500, doi:10.1038/ngeo2724, 2016.